Access to Healthcare and Colorectal Cancer in Kentucky

Matin Katirai PhD, MPH
West Chester University
Department of Geography and Planning
50 University Aveune, Ruby Jones Hall G1-B
West Chester University, West Chester
PA 19383, USA

Abstract

Objectives: *Colorectal cancer (CRC) is one of the major causes of cancer related deaths in the country. The Commonwealth of Kentucky has higher rates of CRC incidence and mortality compared to estimates of national figures. Access to health care is problematic in Kentucky, where 70% of the states county's are designated as rural. The focus of this study is to therefore determine the influence that geographic accessibility to health care and socioeconomic (SES) factors have upon the diagnosis stage of CRC in Kentucky.*

Results: *The variables of race, age, insurance status, median household income and distance to the nearest treatment facility were found to be significant. Results varied by gender.*

Conclusions: *Results indicate that geographic access to health care and SES indicators relate to the stage at diagnosis for CRC patients in Kentucky*

Keywords: Health Care Access, Colorectal Cancer, Geography, Distance, Rural

Introduction

Cancer has placed a tremendous social and economic burden upon our society. According to the Centers for Disease Control and Prevention (CDC), cancer is the leading cause of death among Americans and accounts for approximately 25% of deaths in the United States (U.S.) each year (CDC, 2005). The American Cancer Society (2010) statistics indicate that colorectal cancer (CRC) is a major cause of cancer-related deaths and it is estimated that 51,000 people in the U.S. died in 2010 as a result or CRC. The Commonwealth of Kentucky has higher rates of CRC incidence and mortality compared to estimates of national figures (American Cancer Society, 2010). Kentucky CRC incidence rates for 2002-2006 were approximately 9% higher for men and 6% higher for women compared to estimated U.S. rates. Kentucky CRC mortality rates for were also higher for men and women when compared to estimated U.S. rates.

According to the American Cancer Society, the main risk factor for CRC is age. Other risk factors include family history of CRC or polyps, smoking, obesity, alcohol intake, physical inactivity, and a diet with low consumption of fruits and vegetables but high in fat and/or red meat (Brackley, 2004). Kentucky-focused studies indicate higher prevalences of these risk factors here, compared to other states (Hopenhayn et al.,2004). Stage of disease at diagnosis directly affects mortality for people with CRC, therefore it is an important factor to consider. Persons diagnosed with advanced stage CRC have an estimated 5 year survival rate of 7% (Sheikh et al., 2004). Cancer screening provides an opportunity for detection of early stage cancer, which will dramatically improve the chance for survival. Previous studies have found that areas of residence and socio-economic status (SES) are related to survival, treatment and stage of disease at diagnosis (Brackley & Penning, 2004).

Geographic access to health care is one factor to consider when monitoring a disease like cancer, as geographic location does provide a basis for understanding differences in health outcomes (Schrijvers, Coebergh & Mackenbach, 1997). Geographic access is the distance that must be traveled in order to use health care services, and this can create both cost barriers and time inconveniences (Roderick, Martin & Barnett, 2004). Access to health care is particularly problematic in Kentucky, where 70% of the counties are designated as rural and 42% of the state's counties are in Appalachia, an economically and medically underserved area (Lengerich et al., 2004). Studies in Europe and the U.S. have found that rural patients typically have a more advanced stage of cancer at the time of diagnosis (Campbell et al., 2001). The most disadvantaged rural individuals are those who do not have access to transportation or are members of one-car households, and often include the poor, the elderly, and women (Pickle & Su, 2002). Delivery of rural health care is especially difficult, because individuals who reside in rural areas face other challenges such as limited access to physicians, even when transportation is available (Lyckholm, Hackney & Smith, 2001).

SES considers race, gender, education and economics. Disease outcomes, including those for cancer, are influenced by SES. A clear relationship has been found between patient cancer survival rates and higher SES (Schrijvers, Coebergh & Mackenbach, 1997). Studies have examined SES and found that it is an important predictor of stage at diagnosis, accounting for much of the disparity between whites and non-whites for CRC (Schwartz et al., 2003). Medicaid beneficiaries are 2 to 5 times more likely, than non-Medicaid enrollees to be diagnosed with a late-stage cancer (Bradley, Given & Roberts, 2003). Additionally, people who lack health insurance coverage typically have lower utilization of preventative services, such as screening, further contributing to late stage at diagnosis (Pickle & Yu, 2002). Geographic variations have been found in health insurance coverage, and are useful for understanding patterns of disease outcomes (Winawer, 2005).

Many studies have utilized Geographic Information Systems (GIS) to map rates of cancer incidence and mortality. Other studies have used GIS to correlate proximity to environmental hazards with the development of diseases, such as cancer (Brody et al. 2002). Recently, GIS has been utilized to analyze access to health care facilities. For example, a study by Rushton, Banerjee & Smith (2004) utilized GIS to determine the travel distance required to access health care in Iowa and concluded that local providers of health care diagnose most late-stage CRC cases. The goal of this study was to determine the influence of geographic accessibility to health care and socioeconomic factors have upon the diagnostic stage of CRC in Kentucky. Accessibility to health care is defined, for the purpose of this study, by two measures: the distance from the patient's zip code of origin to the nearest cancer treatment facility and the number of doctors per 1,000 population. The distance from the patient's zip code of origin to the nearest treatment facility was determined with GIS while the physician density was determined from a list of licensed physicians in Kentucky and census population figures. SES factors for this study include: median household income (an aggregate value, zip code level), insurance status of the patient, race, gender, and age. Data about education, an important component of SES, was not available for consideration in this study.

Methodology

Data about incident cases of invasive CRC in the Commonwealth of Kentucky between 1995 and 2002 were retrieved from the Kentucky Cancer Registry (KCR). The data provided by the KCR include patient Identification Numbers (ID) numbers, gender, race, age, zip code and county of patient origin, stage at diagnosis, date of diagnosis, and primary payer status. A list of hospitals was obtained from the KCR and cross-referenced with a list of hospitals that performed CRC procedures obtained from a database provided by the Kentucky Department of Public Health, Division of Epidemiology. The hospital names were then matched with addresses exported from the Kentucky Hospital Association's website. These addresses were used in GIS to determine the distance from each patient's zip code to the nearest hospital. A list of all licensed doctors for 2004 was obtained from the Kentucky Department of Public Health. The zip code of each doctor's practice was included in the data, which allowed the determination of the number of doctors per 1,000 population by zip code.

Median household income, at the zip code level, and total population counts per zip code were retrieved from the U.S. Census Bureau for the year 2000. Zip code points and areas used for GIS analysis were obtained from the Environmental Systems Research Institute (ESRI) GIS data set. Zip code points were used as proxies for the patients' home addresses in the determination of the nearest cancer treatment facilities. Zip code areas are polygons formed from the zip code boundaries. Zip code areas were used for maps concerning late stage at diagnosis. Stage of disease at diagnosis was the dependent variable for the analysis, as defined by the SEER summary Staging Manual. Summary staging is a method of categorizing how far a cancer tumor has spread from its point of origin (National Cancer Institute, 2000). Stages 1-5 and 7 were included in the study, while stage 0, a non-invasive tumor (in situ), and stage 9, an unclassified tumor, were excluded from the analysis. The data included a total of 17,982 cases that were diagnosed in stages 1 – 5 and 7. Individuals with a stage at diagnosis of 2-5 or 7 were coded as 1 and classified as having late-stage diagnosis, as shown in Table 1. Stages 2-5 are classified as regional by the SEER summary manual and signify a tumor extension beyond the limits of the organ of origin. Stage 7 is classified as distant and signifies tumor cells that have traveled to other parts of the body and begun to grow at the new location (metastasized). Those patients who had a stage 1 tumor at diagnosis, which is characterized as local to the organ of origin, were coded as 0 and classified as having early-stage diagnosis. Once stages 0 and 9 had been removed from the KCR data there were a total of 17,982 remaining patient records. Of these, 1,147 were associated with the 550 patients who had multiple primary CRC tumors, and each was classified as a separate case (KCR defines a case by the unique identity of the cancer, not the patient. These were excluded these from the analysis, as the intent of this study was to examine individuals and not cases.

This left 16,835 eligible patients. Table 2 demonstrates that race, insurance status, and median household income variables were similarly distributed within the two groups. Gender, age, distance and the number of doctors per 1,000 population were differently distributed within the two groups. There were a 55.09% males in the multiple case patient group compared to 48.76% in the single case group. There was also a higher percentage of individuals over the age of 70 among the multiple case patients. The proportion of multiple case patients living in areas with greater than 2 doctors per 1,000 population was considerably higher than the proportion for the single case patient group (40.00% vs. 28.19%).

Patient records with incomplete information were also excluded from the study, such as those with an unspecified value of race or insurance status, leaving 16,188 records. In addition, to determine which zip codes were credible, zip codes retrieved from the KCR were compared to a list of zip codes from the U.S. Census. Ninety-five zip codes were not present in the census data. As a further check, these 95 zip codes were cross referenced with the U.S. postal service zip code listing and all 95 zip codes were determined not to be present. The corresponding records were then deleted, leaving a total of 15,705 records remaining or 93.28% of the original number of eligible for the study (Table 3).

Insurance type was one of the measures used for determining SES. Insurance status was determined by the primary payer variable within the KCR data set. There are 15 different insurance categories for the primary payer variable within the KCR data. Only a total of 11 of the 15 categories occurred within the data used for this study (Table 4). The patients were dichotomized into two groups. The first category, coded as 1, consisted of 3 subgroups and included those patients who did not have insurance, those who did not have insurance and were self-pay, and finally those who used Medicaid.Medicaid is a state run, federally funded medical insurance program for the indigent. This first category was considered to represent generally low SES or low access individuals.

All other kinds of primary payer, including insurance, managed care, Medicare, Medicare with supplement, Medicare with Medicare supplement, Tricare, Military, and Veterans Affairs, were grouped to form a higher SES category or higher access, and coded as 0. Insurance status was an individual level variable. Distance between a patient's zip code of origin and the nearest health care facility, median household income, the number of doctors per 1000 population and race were made into dichotomous variables. For distance, a patient with a distance greater than or equal to 10 miles was coded as 1, and a patient with a distance less than 10 miles was coded as 0. For median household income, a patient was coded as 1 if the median household income for the zip code was below $30,000 per year. Median household income was a group level variable. The number of doctors per 1,000 population was coded as 1 if the value was below 2 doctors per 1000 people (Kentucky averages 2.2 doctors per 1000 population[1]). This was a group level variable.

Race was categorized into one of two possible groups. The first group consisted of whites and was coded as 0. All races other than white were categorized into the other group, as the numbers were too small to permit a finer breakdown. The latter group was coded as 1. This was an individual level variable. To determine the distance between a person's zip code of origin and the nearest health care facility, ESRI's ArcView GIS was used in a spatial data analysis. The hospital list from the KCR and the Kentucky Department of Public Health was then geocoded with GIS. A total of 110 hospitals were used in this study to determine geographic access to health care. Of the 110 hospitals, 77 hospitals (70%) were geocoded to a street centerline file. The remaining 33 hospitals that were not located by the geocoding service were manually located and displayed on the map. These 33 hospitals were located in rural areas of the state, where the street centerlines are not fully attributed and are missing street name and number information. The remaining hospitals were located using the hospital address with an internet mapping application (Map Quest).

The GIS map and internet map were then reconciled so that a point could be placed in GIS for the actual hospital location. This was a labor intensive process but it ensured that each hospital was geocoded appropriately. This was preferable to having to omit 30% of the hospitals, which would have led to a substantial reduction in sample size and possibly introduced a sizeable amount of bias. Distances between each patient's zip code and the nearest treatment facility were determined using the ESRI's network analyst in ArcView, which has the capability of determining the nearest facility from any given selected starting point based on a network of street centerlines. The distance from the patient zip code to the nearest treatment facility was calculated individually for all patient zip codes (823 total zip codes).

This process required a starting point, the patient's zip code of origin (zip code centroid), a street centerline file, and the locations of the treatment facilities. The resultant distance obtained from this process was the shortest street length distance in miles from the patient's zip code of origin to the nearest treatment facility. This was a group level variable. Logistic regression analyses were performed to determine the relationship between geographical factors, SES, and stage of disease at diagnosis. The data were first entered into a logistic regression model with both male and female patients included. The data were then stratified by gender and entered into the logistic regression models separately. The outcomes of the logistic regressions included odds ratio estimates (OR), 95% confidence intervals (C.I.), and p-values. Statistical analysis was also performed to describe the distribution of the data. Data were maintained and manipulated in a Microsoft Access database. Statistical analysis of the data was performed in Statistical Analysis System version 8.2 (SAS).

RESULTS

The patient records from the KCR were divided into groups of early and late stage at diagnosis within strata determined by each main covariate considered in the analysis. The distribution of early vs. late stage at diagnosis varied across different strata. Overall 58.61% of patients had a late stage at diagnosis (Table 5). The data reflect that patients were fairly evenly distributed by gender with 48.7% males. The majority of patients diagnosed with CRC were white, accounting for 93.3% of the total cases. Of the total 15,705 patients, 24.9% were over 10 miles away from a cancer treatment facility. The table in Appendix 3 shows more details on the distribution of distance. The majority of patients (58%) were within a five mile distance of the nearest health care facility. Patients that lived in zip code areas that had a median household income of less than $30,000 per year accounted for 36.4% of the data. Patients that lived in zip codes with less than 2 doctors per 1,000 population accounted for 71.81% of the data. Overall of 6.2% of people had no insurance coverage or belonged to Medicaid.

In terms of age distribution approximately 52% of the patients were diagnosed at 70 years of age or older (Table 5). Those in the age group of 60 to 69 constituted about 24% of the total cases (Table 5). Those between 70 and 79 years of age accounted for 31% of the total cases (Table 5). Individuals between 80 and 89 years of age accounted for 18% of the total cases of CRC (Table 5) Table 6 presents a breakdown of the data by age and stage at diagnosis stratified by gender. Both genders seem to follow similar patterns initially. There are higher proportions of late stage cases diagnosed among the younger groups. The difference between males and females begins to emerge in the 40-49 and 50-59 age groups. There is a reduction of approximately 5 % (absolute) in the percentage of males that are diagnosed at a late stage between those two age groups. A similar reduction takes place for women later (between 50-59 and 60-69). Women in the older age groups were more often diagnosed at a late stage compared to men. There are also greater proportions of women in the older age groups diagnosed with CRC compared to men. Women over the age of 80 constituted 26% of all female cases, while men over the age of 80 accounted for 16% of all male cases.

Table 7 displays the results of the logistic regression analyses concerning the association between access to health care, SES and stage at diagnosis (early vs. late). With the full data set the variables of race, age, insurance status, median household income and distance to the nearest treatment facility were found to be significant. When an analysis was performed separately for each gender, distance, age and insurance status were significant for men, while race, median household income and insurance status were significant for women. The analysis of the full data set indicated that individuals coded as non-white for race were at increased risk of being diagnosed at a later stage (OR= 1.17, 95 % C.I. 1.03-1.34), after adjusting for age, insurance status, distance, and median household income. The effect of race appeared to be somewhat stronger among females (OR =1.24, 95% C.I. 1.04-1.48) adjusting for median household income and insurance status. The effect of age was significant with the full data set (OR =0.996, 95% C.I. .993- .998), but this was not so for females (Table 7). Insurance status was the most significant explanatory variable in all three models. For the full data set the OR was 1.50 (C.I. 1.30-1.74) adjusting for race, age, median household income and distance. The effect appeared slightly stronger for males.

Distance was significant with the full data set (OR =1.09, 95% C.I. 1.01-1.18) adjusting for the other variables, however significance was retained only in the model for men, when the genders were considered separately. Then the effect was stronger (OR =1.21). Age is the main risk factor for CRC, but in the logistic regression analysis increased age had a small protective effect for men regarding the stage at diagnosis (OR=.993). This would suggest that for a one year difference in age between two patients holding all other variables constant, the odds of being diagnosed with a late stage of CRC would be only slightly less for the patient who is one year older.

However, the differences are amplified when we consider larger discrepancies in age between two patients holding all other variables constant. Median household income was significant (but not highly so) with the full data set (OR = 1.09, 95% C.I. 1.01-1.16) adjusting for race, age, insurance status and distance. In the gender-specific analyses, this was so only for females.

Discussion

Geographic access was a factor that was found to be significant for men but not women. Men who lived greater than 10 miles away from a health care facility had odds approximately 21% larger of being diagnosed at a late stage for CRC than otherwise similar men living closer to a health care facility. One possible reason for this could be that men living in rural areas distant from health care facilities fail to get preventative screening. Results from Kentucky's Behavioral Risk Factor Surveillance System 2002 Report indicate that 60.7% of men (over the age of 50) have never had a sigmoidoscopy or colonoscopy compared to 52.5% of women (Kentucky Cabinet for Health and Family Service Department for Public Health Division of Epidemiology and Health Planning Surveillance and Health Data Branch, 2004). Studies for prostate cancer have found that men who live in non-urban areas have a higher incidence of late stage of disease at diagnosis and that their screening rates are lower (Jemal et al., 2005). Rural health care facilities have also been found to also provide fewer resource intensive and invasive screening tests such as colonoscopies (Ko, Kreuter & Bladwin, 2005).

Another possible reason for why distance to treatment facility was found to be significant for men but not for women is that women tend to have more visits to health care providers than men. Studies that have examined gender and health care utilization have found that women tend to make use of health care services much more than men (Keene & Li, 2005). Having more visits provides greater opportunity for early detection of disease. In a future study it would be interesting to compare the previous rates of visits to health care providers for the two genders following a diagnosis of CRC. Insurance status was the most significant predictor of stage of disease at diagnosis. Kentucky CRC patients who did not have health insurance coverage or belonged to the state Medicaid program had odds about 50% larger of having a later stage at diagnosis compared to other patients who did have some type of insurance adjusting for the other variables. This is consistent with the findings of other studies that have examined associations between health insurance coverage and stage at diagnosis for cancer (Schwartz et al, 2003; Bradley, Given & Roberts, 2003).

Age is the main risk factor for CRC, but in the logistic regression analysis increased age had a small protective effect for men regarding the stage at diagnosis (OR=.993). A younger male was more likely than an older male to be diagnosed at a late stage of CRC. This could relate to the fact that regular screening is recommended for CRC beginning at the age of 50 (Winawer, 2005). Many people under the age of 50 are not regularly screened for cancer and when signs begin to manifest themselves the disease has usually progressed to a later stage. Age was a significant variable for men and not for women. Studies have found that age-related disparities occur for CRC screening, especially with elderly women (over the age of 80) (Jerent et al., 2005). This may be reflected in our data since there were a greater number (and a higher percentage) of elderly women diagnosed at a later stage.

Race was found to be a statistically significant variable in the female model. Women who were of a race other than white had odds approximately 24% larger of having a late stage at diagnosis. Minority women tend to be the most marginalized in society and are often in the lowest SES group. Socially marginalized groups of people are more likely to lack health insurance coverage. According to U.S. Census figures, 19.4% of African Americans and 32.7% of Hispanics were uninsured compared to 14.7% of non-Hispanic whites (US Census Bureau 2003). Middle-aged African American women are also known to have the highest incidence and mortality rates for cervical cancer and the lowest levels of screening (Hoyo et al., 2005). African Americans and Hispanics have also been found to have lower utilization rates for invasive screening and the fecal occult blood test for CRC (Kentucky Cabinet for Health and Family Service Department for Public Health Division of Epidemiology and Health Planning Surveillance and Health Data Branch, 2004). That both groups have higher rates of no health insurance coverage and lower screening rates helps explain why they may be more likely to be diagnosed at a later stage of disease.

Median household income was found to be significant for women. Findings from this study indicate that women whose household income is less than $30,000 a year, compared to women whose household income is more than $30,000 a year, holding all other variables constant, has odds about 10.5% greater of having a later stage at diagnosis for CRC.

Studies that have examined income level and area of residence were unable to find associations with late stage at diagnosis for other cancers (Keene & Li, 2005). There are several strengths to this study. The first is that there were a large number of patient records, providing a large sample size (15,705). Furthermore, the sample studied accounts for 94.4% of the total eligible population. By eliminating only a small percentage of the eligible population, we minimized the chance for selection bias. The distance that was calculated to the nearest health care facility for this study was a road network distance. A road network distance is more accurate than a straight line distance, which is the shortest distance between two points. Many studies that have considered geographic distance have used straight line distances (Jordan et al. 2004). Road networks take topography into consideration while straight line distances do not.

Another strong feature of this study was that all 110 hospitals were located in the GIS and were matched to exact addresses. Thirty percent of the hospitals were not located by the GIS geocoding service. Standard practice for geocoding when the exact address is not found is to use only a zip code; a point for the location of the hospital would be placed on the map at the zip code centroid location. If this were done or if the data were omitted, there would have been a substantial reduction in sample size or a sizeable amount of bias. One final strong point of this study was that patients who had multiple cases of CRC were not considered. The intent of this study was to examine individuals rather than cases of cancer. Many of the individuals who had multiple cases of CRC had, for their multiple cases, similar values for stage at diagnosis and other variables that were used for this study. If these individuals were included in this study their records would have been unduly influential in the analyses and would have biased the study.

There were several limitations to the study that must be considered. Distances obtained in this study are not actual distances from a patient's home to the nearest health care establishment. This is because the zip code of origin was the only level of address information that was available due to patient confidentiality restrictions. The data provided by the KCR did not indicate where patients were diagnosed and received medical treatment. It was hypothesized that patients would typically travel to the closest medical facilities for treatment and diagnosis, so that our distance variable would be a good proxy for access to health care. This is an assumption and may not always be true, patients have been know to travel to Clinic outside of the state. Some patients move to be closer to treatment, so the distance effect may only exist for diagnosis. Kentucky is primarily a rural state with a considerable portion of the state being in the Appalachian Mountains. Traveling through mountainous terrain is not the same as traveling through non-mountainous areas of the state and therefore a more precise measure of geographic access to health care would have been based on drive time. However, the data set did not contain the necessary information to make such calculations. Figures show that late-stage disease did not always occur at the higher frequencies in Appalachia or in the mountains.

Another assumption made for this study is that everyone living in Kentucky sought medical attention within the state including those people that lived close to a neighboring state. In actuality for some people there may have been closer health care facilities outside of the state, however only Kentucky's health care facilities were considered for this study. Dollar amounts that were used for median household income were based on demographics for entire zip codes. Patients living in the same zip code were assigned the same value for median household income. This implicitly assumes that all patients living in a zip code share similar economic characteristics. However, such an assumption does not hold for all areas, as some zip codes are very large and diverse, and assigning all patients within a zip code the same value is not accurate. A more precise measure would have entailed going to a more detailed level of geography such as census tract or census block group. These areas are at lower levels of geography and would have provided sufficient data for analysis. People living in the same block group or census tract tend to have similar SES. The best information would have been individual level data for household income. However, such information was not available.

Several patients had multiple cases of CRC within the KCR data. Patients with multiple cases of CRC were removed from the sample. Attempts were made to isolate a single case per single patient ID. This involved examining the dates of diagnosis and selecting the earliest date. However, an analysis of the KCR data indicated that for the majority of such patients, the dates of diagnosis were identical, which made it impossible to isolate only one record per patient ID. In the comparison of individuals with multiple cases vs. one case of CRC, there were no dramatic differences between the two groups. If anything it seems that the patients with multiple cases had better indicators which may partly explain why they were diagnosed more closely (i.e. had smaller distances and more doctors per 1,000 population)

The insurance status variable was dichotomized so that it would be an indicator of low or high access to health care. There may be people who were placed into a category to which they did not actually belong. For example, people who had Medicare or military as the primary payer were classified as high access. Many of these people could in reality have been low access that could be potential for misclassification. The list of doctors that was used to define this variable was from 2005. A more accurate variable could have been defined from a list based on the period of 1995-2002 to coincide with the KCR data. The results indicate that geographic access to health care and SES indicators relate to stage of disease at diagnosis for CRC patients in Kentucky. This could be an argument for improving access to health care. Features of improved access to health care would include some type of health insurance coverage and greater access to health care facilities for individuals living in remote locations. By providing some type of health insurance coverage along with improving access to health care for those living in remote locations, we could reduce the frequency of late-stage diagnosis thus increase the survival and quality of life of CRC patients.

Table 1 - Stage classification and distribution of cases

Stage	Stage Classification	# of Records	Recoded
1	Localized	6500	0
2	Regional - by direct extension only	2097	1
3	Regional - lymph nodes involved only	1736	1
4	Regional - by both direct extension and lymph node	1674	1
5	Regional, not otherwise specified	678	1
7	Distant	3020	1

Table 2 - Characteristics of single and multiple case patients

Group of Patients	Male	Non White	Medicaid or no insurance	>=age 70	income >=$30,000	>=10 Miles	>=2 Dr/1000 population
Multiple-case	55.09%	6.56%	5.40%	58.21%	63.94%	20.53%	40.00%
Single-case	48.67%	6.67%	6.16%	51.75%	63.58%	24.85%	28.19%

Table 3 - Determination of study population

Step	Total Records	Percent
File from Cancer Registry	21508	
1-(subtract stage 0 and stage 9)	17982	
2 - Eligible Population	16835	100%
3- Variables with unspecified values removed.	16188	96.16%
4-(valid zip codes) Study Population	15705	93.28%

Table 4 - Primary payer classification and distribution of cases

Payer Code	Payer Classification	# of Cases	Recoded
1	Not Insured	110	1
2	Not Insured, self-pay	296	1
10	Insurance	2618	0
20	Managed Care-HMO, PPO	1557	0
31	Medicaid	562	0
50	Medicare	3091	1
51	Medicare with supplement	7047	0
52	Medicare with Medicare supplement	17	0
53	Tricare	54	0
54	Military	7	0
55	Veterans Affairs	346	0
99	Unspecified (not included in study)	1296	n/a

Table 5 - Characteristics of cases included in study

Stage at Diagnosis		
0-early	Frequency	Percent
1-late	6500	41.39%
	9205	58.61%
Gender		
0-female		
1-male	8061	51.33%
	7644	48.67%
Race		
0-white		
1-other	14658	93.33%
	1047	6.67%
Age Group		
14-29		
30-39	63	0.40%
40-49	310	1.97%
50-59	1091	6.95%
60-69	2370	15.09%
70-79	3743	23.83%
80-89	4795	30.53%
>=90	2901	18.47%
	432	2.75%
Distance		
0-less than or equal to 9.99 miles		
1-greater than or equal to 10 miles	11802	75.15%
	3903	24.85%
Median Household Income, 1999		
0-above $30,000 /year		
1-below $30,000 /year	9986	63.58%
	5719	36.42%
Doctors per 1000 population		
0-greater than or equal to 2 dr/1000 population		
1-less than 2 dr/1000 population	4427	28.19%
	11278	71.81%
INSURANCE STATUS		
0-Insurance or Medicare or other		
1-No Insurance or Medicaid	14737	93.84%
	968	6.16%

Table 6 Distribution of Age Group and Stage at Diagnosis Stratified by Gender

Male Age Group	Early Stage Diagnosis	% early	Late Stage Diagnosis	% late	Total # of Records	% of Males Records	Cumulative
<=29	6	20.69%	23	79.31%	29	0.38%	0.38%
30-39	53	35.33%	97	64.67%	150	1.96%	2.34%
40-49	178	34.36%	340	65.64%	518	6.78%	9.12%
50-59	510	39.44%	783	60.56%	1293	16.92%	26.03%
60-69	853	42.21%	1168	57.79%	2021	26.44%	52.47%
70-79	1033	43.42%	1346	56.58%	2379	31.12%	83.59%
80-89	504	44.84%	620	55.16%	1124	14.70%	98.30%
>=90	60	46.15%	70	53.85%	130	1.70%	100.00%
Female Age Group	**Early Stage Diagnosis**	**% early**	**Late Stage Diagnosis**	**% late**	**Total # of Records**	**% of Females Records**	**Cumulative**
<=29	11	32.35%	23	67.65%	34	0.42%	0.42%
30-39	68	42.50%	92	57.50%	160	1.98%	2.41%
40-49	217	37.87%	356	62.13%	573	7.11%	9.51%
50-59	397	36.86%	680	63.14%	1077	13.36%	22.88%
60-69	711	41.29%	1011	58.71%	1722	21.36%	44.24%
70-79	1049	43.42%	1367	56.58%	2416	29.97%	74.21%
80-89	734	41.31%	1043	58.69%	1777	22.04%	96.25%
>=90	116	38.41%	186	61.59%	302	3.75%	100.00%

Table 7: Odds Ratio Estimates, 95% Confidence Intervals and P Values

Model – 1 All Records			
Variable	**Odds Ratio**	**P Value**	**95 % Confidence Interval**
RACE	1.17	0.0148	1.032 to 1.341
AGE	0.996	0.0016	0.993 to 0.998
INS (Insurance Status)	1.50	<.0001	1.296 to 1.738
MINC (Median Household Income)	1.086	0.0177	1.014 to 1.163
Distance to health care	1.094	0.021	1.014 to 1.180
Model - 2 Male			
Variables	**Odds Ratio**	**P Value**	**95 % Confidence Interval**
Age	0.993	0.0001	.989 to .996
INS	1.604	<.0001	1.289 to 1.996
DISTANCE	1.207	0.0005	1.085 to 1.341
Model - 3 Female			
Variables	**Odds Ratio**	**P Value**	**95 % Confidence Interval**
RACE	1.242	0.0167	1.040 to 1.483
INS	1.473	<.0001	1.218 to 1.782
MINC	1.105	0.0363	1.006 to 1.212

Refrences

American Cancer Society (ACS). (2010). Cancer Facts & Figures. http://www.cancer.org/downloads/STT/CAFF_finalPWSecured.pdf. Accessed 05/13/2011.

Brackley ME, Penning MJ. (2004).Residence, income and cancer hospitalization in British Columbia during a decade of policy change. *International Journal for Equity in Health*, 15;3(1):2.

Bradley CJ, Given CW, Roberts C. (2003). Late stage cancers in a Medicaid-insured population. *Medical Care*. 41(6):722-8.

Brody JG, Aschengrau A, McKelvey W, Rudel RA, Swartz CH, Kennedy T. (2004). Breast cancer risk and historical exposure to pesticides from wide-area applications assessed with GIS. *Environmental Health Perspective*.112(8):889-97.

Brody JG, Vorhees DJ, Melly SJ, Swedis SR, Drivas PJ, Rudel RA. (2002). Using GIS and historical records to reconstruct residential exposure to large-scale pesticide application. *Journal of Exposure Annals Environmental Epidemiology*.12(1):64-80.

Campbell NC, Elliott AM, Sharp L, Ritchie LD, Cassidy J, Little J. (2001). Rural and Urban Differences in Stage at diagnosis of colorectal and lung cancers. *British Journal of Cancer*. 84(7),910-14.

Centers for Disease Control. (2009)Colorectal Cancer Fact Sheet. Colorectal Cancer:

http://www.cdc.gov/cancer/colorectal/. Accessed 08/27/2009/

Hopenhayn, C., Moore, D. B., Huang, B., Redmond, J., Tucker, TC., Kryscio, R. J., Boissonneault, G.A. . (2004). Patterns of Colorectal Cancer Incidence, Risk Factors, and Screening in Kentucky. *The Southern Medical Journal*. 97(3):216-23.

Hoyo C, Yarnall KS, Skinner CS, Moorman PG, Sellers D, Reid L.(2005). Pain predicts non-adherence to pap smear screening among middle-aged African American women. *Preventative Medicine*. 41(2):439-45.

Jemal A, Ward E, Wu X, Martin HJ, McLaughlin CC, Thun MJ.(2005). Geographic patterns of prostate cancer mortality and variations in access to medical care in the United States. *Cancer Epidemiology Biomarkers and Prevention*. 14(3):590-5.

Jerent AF, Franks P, Jackson JE, Doescher MP. (2004). Age-related disparities in Cancer Screening. *Annals of Family Medicine*.2(5)481-7.

Jordan H, Roderick P, Martin D, Barnett S. (2004).Distance, rurality and the need for care: access to health services in South West England. *International Journal of Health Geographic's*. Sep 29;3(1):21

Keene J, Li X. (2005). Age and gender differences in health service utilization. *Journal of Public Health*. 27(1):74-9.

Kentucky Cabinet for Health and Family Service Department for Public Health Division of Epidemiology and Health Planning Surveillance and Health Data Branch. (2004) *Kentucky Behavioral Risk Factor Surveillance System 2002 Report*. Frankfort, Kentucky.

Kennedy H.(2001), *Dictionary of GIS Terminology*. Redlands, CA: ESRI Press.

National Cancer Institute. SEER Summary Staging Manual (2005). Available http://seer.cancer.gov/tools/ssm/ **.** Accessed 06/06/2005.

Ko WC, Kreuter W, Baldwin LM. (2005)Persistent demographic differences in colorectal cancer screening utilization despite Medicare reimbursement. *BioMedCentral Gastroenterology*. 10:1471-1479.

Lengerich, E J.,Tucker, T.C., Powell, R. K.,Colsher, P, Lehman, E., Ward, A. J. , Siedlecki, J. C., Wyatt, S.W. (2005). The Journal of Rural Health. 21(3):39-48.

Lengerich EJ, Wyatt SW, Rubio, S., Beaulieu, JE., Coyne, C, Fleisher, L., Ward, AJ. (2004). The Appalachia Cancer Network: cancer control research among a rural, medically underserved population. The *Journal of Rural Health***.** Spring;20(2):181-7.

Lyckholm LJ, Hackney MH, Smith TJ. **(**2001) Critical Reviews in Oncology/Hematology. *Ethics of Rural Health Care*. 40(2):131-8.

Pickle LW, Su Y. (2002). Within-State Geographic Patterns of Health Insurance Coverage and Health Risk Factors in the United States. *American Journal of Preventative Medicine*. 2:75-83.

Rushton G., Peleg I., Banerjee A., Smith G., West M. (2004). Analyzing Geographic Patterns of Disease Incidence: Rates of Late-Stage Colorectal Cancer in Iowa. *Journal of Medical Systems*. 28(3):223-36.

Schrijvers T.M, Coebergh JW, Mackenbach JP. (1997).Socioeconomic Status and Comorbidity among Newly Diagnosed Cancer Patients. *Cancer*. 80(8):1482-8.

Schwartz KL, Crossley-May H, Vigneau FD, Brown K, Banerjee M. (2003)**.** Race, socioeconomic status and stage at diagnosis for five common malignancies. *Cancer Causes and Control*. 14(8):761-6.

Sheikh RA, Kapre S, Calof OM, Ward C, Raina A. (2004). Screening Preferences for Colorectal Cancer: A Patient Demographic Study. *The Southern Medical Journal*. March; (3):224-230.

US Census Bureau. Health Insurance Coverage: (2003). Available :

http://www.census.gov/hhes/www/hlthins/hlthin03/hi03t5.pdf Accessed 06/13/2005/

Winawer SJ. (2005) Screening of colorectal cancer: progress and problems. *Recent Results Cancer Research*. 166:231-44.